

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 18, 2003, 15:04:29 ; Search time 15.0398 Seconds

(without alignments)  
2160.495 Million cell updates/sec

Title: US-09-807-933B-1

Perfect score: 1836

Sequence: 1 MKFTTASALLALALGTEM.....TYKEVTCPEKITAKTGCSSRK.338

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	708.5	38.6	229	2 UC7308	cellulase (EC 3.2.
2	552.5	30.1	511	2 S10527	endoglucanase B pr
3	515.5	28.1	393	2 S59499	cellulase egli - s
4	241	13.1	471	1 A26160	cellulose 1,4-beta
5	241	13.1	471	1 A38979	cellulose 1,4-beta
6	188.5	10.3	410	1 S68153	cellulase (EC 3.2.
7	180	9.8	418	1 S28372	cellulase (EC 3.2.
8	173	9.4	438	1 S70802	cellulose 1,4-beta
9	153.5	8.4	1428	2 T08852	luciferin A - Califo
10	150	8.2	2704	2 S09118	G surface protein
11	149.5	8.1	475	2 S49886	probable membrane
12	147.5	8.0	2395	1 S50820	surface protein ty
13	147	8.0	242	2 S60143	cellulase (EC 3.2.
14	144	7.8	2718	2 A23475	G surface protein
15	143	7.8	1396	2 S36851	L-shaped tail fibre
16	142	7.7	860	2 S43846	xylinase B - rumen
17	141	7.7	962	2 S03818	carboxymethylcellu
18	138	7.5	495	2 B71360	hypothetical prote
19	137	7.5	1217	2 S52714	sericinB - silkwo
20	137	7.5	4776	2 B95206	cell wall surface
21	136.5	7.4	388	1 UC5461	cellulase (EC 3.2.
22	136	7.4	1275	2 T33369	hypothetical prote
23	135.5	7.4	881	2 S56032	probable membrane
24	135	7.4	302	2 S71334	acetyl xylan ester
25	135	7.4	388	2 S43920	cellulase (EC 3.2.
26	135	7.4	888	2 T46726	secreted acid phos
27	133	7.2	513	1 EUT01	cellulose 1,4-beta
28	133	7.2	513	2 S45380	cellulose 1,4-beta
29	132	7.2	286	2 C61615	sericin MG-2 - gre

30	132	7.2	513	2 S11439	cellulose 1,4-beta
31	130.5	7.1	1063	2 D86731	hypothetical prote
32	129	7.0	316	1 A38743	luciferin - human
33	128.5	7.0	1560	2 T02885	peroxisome prolif
34	127.5	6.9	537	2 B33485	spore coat protein
35	126.5	6.9	604	1 JP0001	glucan 1,4-alpha-g
36	126.5	6.9	856	2 T00349	Avicelase III - As
37	125.5	6.8	1217	2 T25894	hypothetical prote
38	125.5	6.8	2403	2 A53386	sauro human
39	125	6.8	477	2 S53362	mucin 5AC (clone J
40	124.5	6.8	420	2 S53916	SUN4 protein precu
41	124.5	6.8	459	2 A25928	cellulase (EC 3.2.
42	124.5	6.8	786	2 T33585	ubiquitin protein
43	124	6.8	344	2 T40167	hypothetical prote
44	124	6.8	1131	2 T41144	hypothetical serin
45	123.5	6.7	402	2 E86185	hypothetical prote

#### ALIGNMENTS

```
RESULT 1
UC7308
cellulase (EC 3.2.1.4) - Scopulariopsis brevicaulis
N/Alternate names: endoglucanase I
C/Species: Scopulariopsis brevicaulis
C/Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
C/Accession: UC7308; PC7087
R/Nakata, F.; Kawaguchi, T.; Takada, G.; Sumitani, J.; Moriyama, Y.; Arai, M.
Biosci. Biotechnol. Biochem. 64, 1238-1246, 2000
A/Title: Cloning and sequencing of an endoglucanase gene from Scopulariopsis brevicaulis
A/Reference number: UC7308
A/Accession: UC7308
A/Molecule type: DNA
A/Residues: 1-229 <NA>
A/Experimental source: strain TOF-1212
A/Accession: PC7087
A/Molecule type: protein
A/Residues: 21-37;149-164 <NA>
C/genetics:
A/gene: eg1
A/introns: 147/3
C/keywords: glycosidase; hydrolase

Query Match      38.6%; Score 708.5; DB 2; Length 229;
Best Local Similarity 57.5%; Pred. No. 9.8e-39;
Matches 122; Conservative 34; Mismatches 49; Indels 7; Gaps 5;

QY      129 ASGNQVTRRYMDCKASCSMPGKANVS-SPVKSNKNDGVTALSDSN-AQSGCN-GGNSYM 185
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      21 ASGTGTTTRYPDCCKPSCSPDKAPLSQGPMTCTDIND-NPLDDGLTSGCEPGGAYM 79

QY      186 CNDNQPMVAVNDNLAVGFAAAISGGESRMCCSELTFTSTSVAGKKVAVVNTGDL 245
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      80 CSHSPMAVDDELAVGMAAVNIGQGTESDMCCACYLEFTTGAVSGKKMIVQATNTGDL 139

QY      246 GSSTGAPFLQMPGCGVGIFNCCSSQMGAPNDGMSRYGIGISASDCLSPALQAGCW 305
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      140 GNN--HFDIAMPGGVGIFNCGTIDQWSPPMGMDRVRGVTTRADCSFPALVAGCBW 196

QY      306 RFNMFADNPMTYKEVTCPEKITAKTGCSSR 337
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      197 RFDFGQIDNPVSRFEVCEPALVOKSCOR 228

RESULT 2
S10527
endoglucanase B precursor - Pseudomonas fluorescens
C/Species: Pseudomonas fluorescens
C/Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 08-Oct-1999
C/Accession: S10527
R/Gilbert, H.J.; Hall, J.; Hazlewood, G.P.; Ferreira, L.M.A.
Mol. Microbiol. 4, 759-767, 1990
```





Db 1237 SSGSGSASGSGTGGKGSAYDTDADSGSDNRSPGYLLFODPCTPGLYINCTAGTCRLTAWC 1236

A:Gene: SGD:SIM1  
A:Cross-references: SGD:S0001385; MIPS:YIL123w

A:Map position: 9L  
 C:Superfamily: Saccharomyces NCA3 protein  
 C:Keywords: transmembrane protein  
 F:82-98/Domain: transmembrane #status predicted <tm>

Query Match 8.1%; Score 149.5; DB 2; Length 475;  
 Best Local Similarity 20.4%; Pred. No. 0.015;  
 Matches 71; Conservative 47; Mismatches 131; Indels 99; Gaps 13;

QY 5 TLASSLLALALGTEMAAECKLYGCGCKMNGPTCCSGSTCKVSNDRYSQCLPBG 64  
 DB 79 TSSAAGIAIASIAVSSAALAKNEKISDAAA---SATASTSQASSSSSSSSATSTLESS 134  
 QY 65 SSGNKSSESAAHKT---TTAAHKTTTAAHKTTPAKKTTVAKASTPNSSSSSSK 121  
 DB 135 SVSSSEEAAPTSTVSTSTSAATOSASATKSTSTSTSTSTSTSTSTSTSTSTSTSSSSSS 194  
 QY 122 YSAVSGA-----SGNGVTRVY----- 139  
 DB 195 SSSSSGSGSIYGLADPFGSPBEKFDGTFPCDKFPGQGVISIDWIGGWSGVENTDTS 254  
 QY 140 ----DC-----CKASCWPGKANVSPLYKSNKDGVTALDSNAQSG---CNGNSYKQNN 189  
 DB 255 TGGSCKEGYSYSCQ--PGMSKTQWPSDQ-----PSDGRSVGLCKNGLYLSNTD 305  
 QY 190 QPVAVDNLAIFYFAAIAISGGSGSRWCCSFELTFTSTSVAG--KKRVVQVNTNGDL--- 245  
 DB 306 ADLCWGVBAIVVSKLSKG-----VAICRTDYPGRNVIPRYEGSSSLP 354  
 QY 246 -----GSGTAHFDLMPGGGVIFNGCSSQCAPNDGMS 281  
 DB 355 TVVDQDTYFTWEGKTSAGY--VYNNAGVSEVDCI--WGTSGSGIGN 398

## RESULT 12

SS0820  
 Surface protein type 51B - Paramedium tetraurelia  
 C:Species: Paramedium tetraurelia  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 07-Dec-1999  
 C:Accession: S50820  
 R:Scott, J.; Leeck, C.; Roney, J.  
 Nucleic Acids Res. 22, 5079-5084, 1994  
 A:Title: Analysis of the micronuclear B type surface protein gene in Paramedium tetraurelia  
 A:Reference number: S50820; MUID:95098630; PMID:7800503  
 A:Accession: S50820  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-2395 <SCD>  
 A:Cross-references: EMBL:U07603; NID:9467226; PID:AAA81947.1; PID:9467227  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1994  
 C:Genetics:  
 A:Genetic code: SGCS  
 A:Introns: 472/3; 1310/3; 1821/3  
 C:Superfamily: G surface protein

Query Match 8.0%; Score 147.5; DB 1; Length 2395;  
 Best Local Similarity 22.1%; Pred. No. 0.091;  
 Matches 83; Conservative 38; Mismatches 149; Indels 105; Gaps 18;

QY 22 SAABCSKLYGCGCKMNGPTCC-----SGSTC-----KVSND-----YYSQC 60  
 DB 1953 SKGACFR-YTSCSKLWNSDPTSCKMISNOCTTNGSNCIGITLCSFNTDQGVSGYDAGC 2011  
 QY 61 LPSGSGNGKSSSEAAHKTAA-----HKTTAAHKTTPAKKTTVAKASTPNS 114  
 DB 2012 IQGVPLNNSDPVPCPYTSCAFAFTTSHDCCOTASKKTTNG-----TTGCTALACGS 2066  
 QY 115 SSSSSSKY-----SAVSGASGNGVTRVWDCCKASC-----SWPGKANVSAPVS 160  
 DB 2067 YTTQACQFYNDKALILSGAITSITGCT--WDTTASSCRQSCADLTGTHAACSSQLST 2124  
 QY 161 CNKQGVTALE---SDSNAGSGNG--GNSYMCNDQPAVVDNLAYG--FAAAISGGG 211

DB 2125 CTSDETSCLVKGACTSYTTQTACTTAVGSDGICVWELASSTNNNTAKRCLLACADIONGT 2184  
 QY 212 ESRWC-----CSCF--ELFTSTSVAGKXWVQVNTNGDSSSTG 250  
 DB 2185 STNVCAVALSCVNSGCTACIACANCSYTTTKACNSGGLDGI CVTQSTATATAAGTGTC 2244  
 QY 251 AHFDLQMPGGVGVIFNGCS-----QMGAPNDGW--GSRVGISSASDCS-----SLP 296  
 DB 2245 A-----LMTACTANSQVACQAAKDRCSWTAASGTGTTAVASKCATHTCATN 2292  
 QY 297 SALQACCKRFRWFK 311  
 DB 2293 QATNGACTRFLNWDK 2307

## RESULT 13

S60143  
 cellulase (EC 3.2.1.4) V precursor - fungus (Trichoderma reesei) (strain QM9414)  
 N:Alternate names: endo-1,4-beta-glucanase V  
 C:Species: Trichoderma reesei  
 A:Variety: strain QM9414  
 C:Date: 15-Feb-1996 #sequence\_revision 01-Mar-1996 #text\_change 17-Mar-1999  
 C:Accession: S60143; S49043  
 R:Salcherno, A.; Henriksen, B.; Hoffren, A.; Telesman, O.; Penttila, M.  
 submitted to the EMBL Data Library, May 1994  
 A:Description: A novel small endoglucanase gene egls from Trichoderma reesei isolated by  
 A:Reference number: S60143  
 A:Accession: S60143  
 A:Molecule type: DNA  
 A:Residues: 1-242 <SAL>  
 A:Cross-references: EMBL:Z33381; NID:9485863; PID:9485864  
 R:Salcherno, A.; Henriksen, B.; Hoffren, A.M.; Telesman, O.; Penttila, M.  
 Mol. Microbiol. 13, 219-228, 1994  
 A:Title: A novel, small endoglucanase gene, egls, from Trichoderma reesei isolated by e  
 A:Reference number: S49043; MUID:95075308; PMID:7984103  
 A:Accession: S49043  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: DNA  
 A:Residues: 206-241 <SAM>  
 A:Cross-references: EMBL:Z33381  
 C:Genetics:  
 A:Gene: egls  
 A:Introns: 23/3; 45/3  
 C:Function:  
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce  
 A:Substrate: cellulose degradation  
 C:Superfamily: fungal cellulose-binding domain homology  
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation  
 F:1-17/Domain: signal sequence #status predicted <SIG>  
 F:18-242/Product: cellulase #status predicted <MAT>  
 F:210-241/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 8.0%; Score 147; DB 2; Length 242;  
 Best Local Similarity 67.6%; Pred. No. 0.011;  
 Matches 23; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 29 LYGGCGCKMNGPTCCSGSTCKVSNDRYSQCLP 62  
 DB 209 LYGGCGAGWNPPTQAPGTCKVQWYSQCLP 242

RESULT 14  
 A23475  
 G surface protein - Paramedium primaurelia  
 C:Species: Paramedium primaurelia  
 C:Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 07-Dec-1999  
 C:Accession: A23475  
 R:Prat, A.; Katinka, M.; Caron, F.; Meyer, E.  
 J. Mol. Biol. 189, 47-60, 1986  
 A:Title: Nucleotide sequence of the Paramedium primaurelia G surface protein. A huge pro  
 A:Reference number: A23475; MUID:87060934; PMID:3783679  
 A:Accession: A23475  
 A:Molecule type: DNA

